Popular Environmental Sampling Designs

1

Sampling designs in common use

These are all probability-based (statistical) and all are well documented in the environmental literature:

- Simple Random
- Systematic/Grid
- Stratified
- Composite
- Incremental
- Rank Set
- Hot spot
- Adaptive Cluster

2

Simple Random Sampling

- For this kind of sampling (SRS) every part of the site or population has an equal chance of being in the sample
- A map of the site or a complete list of the population is constructed and every potential sample point identified
- Using a random process (primitive; numbers-from-a-hat, sophisticated; random number generator) select a fixed number of samples to be collected
- Translate the sample points to be collected from the map or list to the physical site

3

Simple Random Sampling: Pros

- Simple in concept and provides proper data (theoretical support) for statistical data analysis
- Protects against bias in estimating parameters (e.g., means) and testing hypotheses
- Is the basic building block of more complicated (and efficient) sampling designs

4

Simple Random Sampling: Cons

- Ignores available information that could be used to develop more cost-effective sampling designs
- Not as efficient as other designs for delineating patterns of contamination or finding hot spots
- Difficult to find randomly selected sampling locations
- May not be truly representative of the population
- Tends to demand large numbers of samples

5

Systematic (Grid) Sampling

- Systematic (grid) sampling consists of collecting samples according to a specified pattern at regular intervals in space or time
- Samples may be collected at the center of each individual grid space or at the nodes of the grid
- The grid is orientated at random across a site or with a random starting point if along a line
- Examples:
  - Square grid patterns over a site
  - equal-interval sampling along a straight line

6

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **MAR 2012**	2. REPORT TYPE	3. DATES COVERED **00-00-2012 to 00-00-2012**
4. TITLE AND SUBTITLE **Popular Environmental Sampling Designs**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Neptune and Company, Inc,1435 Garrison Street, Suite 110,Denver,CO,80215**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Presented at the 9th Annual DoD Environmental Monitoring and Data Quality (EDMQ) Workshop Held 26-29 March 2012 in La Jolla, CA. U.S. Government or Federal Rights License

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**6**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Systematic Sampling: Pros

- Easy to explain and implement and provides uniform coverage of site or project
- Good for estimating boundaries, trends or patterns of contamination over space or time.
- May yield more precise estimates of population parameters than other sampling designs
- Required for statistical data analysis to estimate trends and spatial patterns
- Easier than Simple Random to locate samples

7

Systematic Sampling: Cons

- Systematic sampling can cause estimated means to be biased if the sampling grid pattern lines up with any pattern of contamination
- More information is needed (than for simple random sampling) about the population to estimate the variance of the estimated mean
- May require a large number of samples if all squares or nodes sampled

8

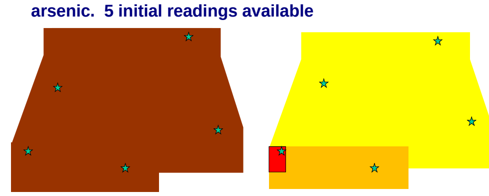
Stratified Sampling

- The target population is divided meaningfully into contiguous sub-populations called strata
- The concept is to divide a relatively high variability population into relatively homogeneous strata
- Sampling locations are selected independently within each stratum using some sampling design (Simple Random, for example)

9

Example: Littlewood Site

- Multi-purpose site is suspected to be contaminated with arsenic. 5 initial readings available



Site before stratification
0.2 – 33.0 ppm
The high variability indicates many samples will be needed for further exploration

Site after stratification
Parking 0.2 – 0.4 ppm
Storage 5.6 ppm
Production 33.0 ppm
More efficient sampling

10

Stratified Sampling: Pros

- Dramatically reduces the variability present in the population and hence improves overall precision
- Enables estimates of individual areas to be made and so builds a better conceptual model
- Assists in providing good coverage of the project by not allowing too many samples in the same area
- Allows for increased samples from sensitive or more important areas (Parking vs Storage vs Production)

11

Stratified Sampling: Cons

- Requires advance knowledge in order to divide the study area into roughly homogeneous strata before sampling
- The number of samples to be taken in each stratum must be determined (Simple Random or other?)
- If strata boundaries inaccurate, what appears to be outlier data can be due to being incorrectly identified to stratum wrong one) and adversely influence conclusions (Parking vs Storage vs Production)

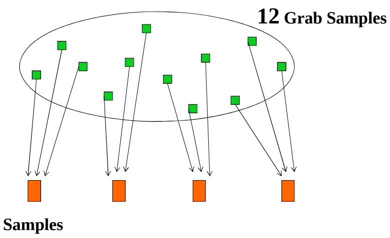
12

Composite Sampling

- Many individual (grab) samples are combined and thoroughly mixed to make a homogeneous whole
- At random, subsamples (composite samples) are made and sent to the laboratory for analysis
- The physical size of composite samples are the same size as those obtained at random

13

Composite Sampling



14

Composite Sampling: Pros

- Allows for estimating the mean contamination with the same precision at a lower cost
- Provides better coverage of the study site without increasing the number of chemical analyses
- Allows for a more representative sample from a basic area of sample support (sampling unit)
- Can be used in combination with other sampling designs

15

Composite Sampling: Cons

- Information on individual samples used to form composite samples is lost in compositing
- Potential for loss of contaminants (volatiles) during the mixing and handling phase
- Potential for reactions and interactions among analytes during compositing
- Need to make decision on how many grab samples to be composited and how many composite samples to send for analysis

16

Incremental Sampling Design

- Similar to composite sampling but covers a defined decision unit
- After deciding on the mass (volume) to be sent for analysis (e.g. 1500 grams), a large number of smaller, incremental, samples are selected (e.g. 30 increments of 50 grams each)
- Necessary to define the QA protocol and technique for taking each increment to prevent bias

17

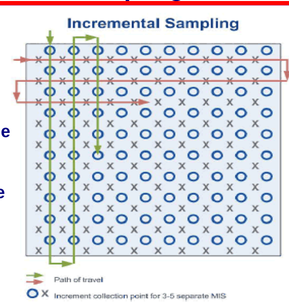
Incremental Sampling

The red path shows the collection of the first incremental sample.

The green path shows the second incremental sample (replicate).

A third incremental sample could be a diagonal path.

Each of these incremental Samples consists of 100 increments



18

Incremental Sampling: Pros

- Provides better coverage of the decision area (unit) without increasing the number of chemical analyses
- Allows for a more representative sample from a basic area of sample support (sampling unit)
- Highly effective when large variability in results suspected (e.g. minor hotspots)
- Estimates from incremental sampling very stable

19

Incremental Sampling: Cons

- No understanding of spatial variability is gained
- Only valid for estimating the mean
- Decision on how many increments to take is arbitrary
- Potential for loss of contaminants (volatiles) during the mixing and handling phase
- Comparison of incremental samples to other samples is difficult
- Size / shape of decision unit not adjustable after sample collection

20

Ranked Set Sampling

- A sampling design where expert judgment is used in combination with simple random sampling
- Decide on how many samples are needed (m)
- Using SRS, create m x m (m-squared) potential samples (note potential samples identified, but not collected)
- From each set of n potential samples, the expert then ranks these potential samples according to some criterion
- One sample is then selected from each set of potential samples to send for analysis

21

Statistician & Expert

- Statistician:
  - Selects "m" sets of random samples of size "m" as potential samples (total "m x m") using SRS
- Expert:
  - Within each set, the expert ranks (grades) the potential samples from highest to lowest based on the expert's opinion or some criterion
- Together:
  - From the first set, the largest is chosen; from the second set, the second largest is chosen; from the third set, the third largest is chosen etc
- Result:
  - A random sample of size "m" with special properties

22

Ranked Set Sampling: pros

- Increased chance of good representativeness through the incorporation of expert advice
- Better precision than Simple Random Sampling that can be most impressive depending on data distribution
- Same simple formulae to use, no special adjustments
- Mistakes in ranking still leaves the resulting sample better than SRS
- Even if the expert advice is useless the resulting sample is as good as SRS

23

Ranked Set Sampling: cons

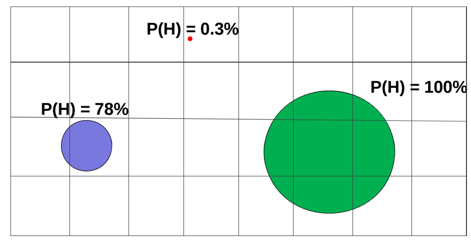
- Need to find some variable to do the ranking on
- Possible increased costs of having to pay the expert
- Difficulty quantifying the exact amount of improvement the resulting "super sample" is over SRS

24

Hot-Spot Sampling

- Detecting the presence of a potential "hot-spot" by covering the site with a grid of known size and mesh-spacing, then taking samples at the nodes of the grid
- Define the area, volume, and shape of the hot spot
- Should estimate the concentration of the potential hot-spot
- Need to decide the risk of missing a hot-spot

Hot-Spot Size and Grid Spacing



Hot-Spot Sampling: Pros

- Very effective in finding hot spots of a specified size and shape
- Able to control the probability of missing a hot-spot of a given size
- Indirectly delineates the size and shape of the hot-spot

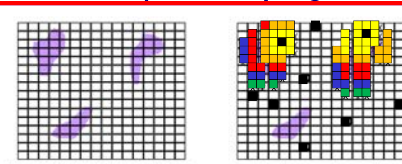
Hot Spot Sampling: Cons

- Need to specify the size and shape the potential hot-spot and this could be difficult
- Need to decide the acceptable chance of missing a hot-spot of a certain size – should be related to risk
- Potentially expensive due to need to take samples from all over the site
- Need the capacity to conduct multiple rounds of sampling

Adaptive Sampling

- Select an initial probability based sample, evaluate the results, then choose additional samples based on results from the initial sample
- The additional samples are taken all around the initial samples having an elevated level of interest (e.g., high contamination level)
- Repeat additional sampling until further samples show nothing of interest (e.g., lower level of contamination)

Adaptive Sampling



Areas with elevated concentrations — Adaptive sampling results

The purple areas represent unknown contamination

Black	*Initial sample (there were 3 hits)*
Yellow	*1^{st} round (there were 6 hits)*
Orange	*2^{nd} round (there were 6 hits)*
Red	*3^{rd} round (there were 8 hits)*
Blue	*4^{th} round (there were 2 hits)*
Green	*5^{th} round (there were no hits)*

Adaptive Sampling: Pros

- A clear advantage in delineating and investigating the size of potential hot-spots
- Allows for the ability to estimate the population mean and variance using all the available data
- Research shows that the variance of the population mean estimate can be considerably smaller than that for simple random sampling leads to better accuracy of results

Adaptive Sampling: Cons

- Complicated computation of statistics such as mean and variance for use in making probabilistic statements
- Time-consuming: iterative sampling required until no more samples of interest are found
- Sample sizes usually much larger than traditional designs causing greater cost

Sampling Designs: Conclusions

- Probabilistic more defensible than judgmental
- If only relatively few samples can be afforded, then more sophisticated designs will be needed
- The more sophisticated the design, the more sophisticated the analysis
- Advice: *Guidance on Choosing a Sampling Design for Environmental Data Collection (G-5S)* download from: www.epa.gov/quality/qa_docs.html